Continous Data Analysis Exercise Solutions

Set-up

```
library(tidyverse) # for tidyverse
library(here) # for file paths
library(survey) # for survey analysis
library(srvyr) # for tidy survey analysis

recs <- read_rds(here("Data", "recs.rds"))

recs_des <- recs |>
  as_survey_rep(weights=NWEIGHT,
                repweights=starts_with("BRRWT"),
                type="Fay",
                rho=0.5,
                mse=TRUE)
```

Part 1

1. Find the average square footage of housing units (TOTSQFT_EN) with a 90% confidence interval.

```
avg_sqci<-recs_des |>
  summarize(
    SF_HU=survey_mean(TOTSQFT_EN,
                      vartype = "ci",
                      level = 0.9)
  )
avg_sqci
```

```
# A tibble: 1 x 3
  SF_HU SF_HU_low SF_HU_upp
  <dbl>     <dbl>     <dbl>
1 2008.     1981.     2036.
```

On average US households have 2,008 square feet, with a 90% CI of (1,980 sq ft, 2,036 sq ft).

2. Estimate the ratio of cooled square footage to total square footage (TOTCSQFT) to the total square footage of housing units (TOTSQFT_EN) with its standard error.

```
cool_totratio<-recs_des |>
  summarize(
    PropCooled=survey_ratio(
      numerator = TOTCSQFT,
      denominator = TOTSQFT_EN,
      vartype = "se")
  )
cool_totratio
```

```
# A tibble: 1 x 2
  PropCooled PropCooled_se
       <dbl>         <dbl>
1      0.685       0.00938
```

On average US households have a ratio of 0.68 square feet cooled per total square feet.

3. Estimate the median temperature housing units are set to during the night in the winter (WinterTempNight) using the `survey_median` function.

```
med_wintertemp<-recs_des |>
  summarize(
    temp_winter=survey_median(WinterTempNight,
                              vartype = "se",
                              na.rm = TRUE)
  )
med_wintertemp
```

```
# A tibble: 1 x 2
  temp_winter temp_winter_se
        <dbl>          <dbl>
1          68          0.252
```

The median temperature housing units are set to during the night in the winter is 68 degrees Fahrenheit.

4. Estimate the median temperature housing units are set to during the night in the winter (WinterTempNight) using the `survey_quantile` function.

```
recs_des |>
  summarize(
    WinterNightTemp=survey_quantile(WinterTempNight,
                                    quantiles = 0.5,
                                    vartype = "se",
                                    na.rm = TRUE)
  )
```

```
# A tibble: 1 x 2
  WinterNightTemp_q50 WinterNightTemp_q50_se
                <dbl>                  <dbl>
1                  68                  0.252
```

The 50th percentile (median) temperature housing units are set to during the night in the winter is 68 degrees Fahrenheit.

Part 2

1. Estimate the total average energy cost (TOTALDOL) by region, division, and urbanicity.

```
# option 1
recs_des |>
  group_by(Region, Division, Urbanicity) |>
  cascade(
    EnergyCost=survey_mean(TOTALDOL)
  )
```

```
# A tibble: 45 x 5
# Groups:   Region, Division [15]
   Region    Division           Urbanicity    EnergyCost EnergyCost_se
   <fct>     <fct>              <fct>              <dbl>         <dbl>
 1 Northeast New England        Urban Area         2629.          95.2
 2 Northeast New England        Urban Cluster      1889.         214.
 3 Northeast New England        Rural              2865.         109.
 4 Northeast New England        <NA>                2546.          61.3
 5 Northeast Middle Atlantic    Urban Area          2133.          32.9
 6 Northeast Middle Atlantic    Urban Cluster       2259.         193. 
 7 Northeast Middle Atlantic    Rural               2420.         108. 
 8 Northeast Middle Atlantic    <NA>                2174.          41.7
 9 Northeast <NA>               <NA>                2273.          34.7
10 Midwest   East North Central Urban Area          1632.          43.2
# i 35 more rows
```

```
# option 2
# one way
recs_des |>
  group_by(Region, Division, Urbanicity) |>
  summarize(
    EnergyCost=survey_mean(TOTALDOL)
  )
```

```
# A tibble: 30 x 5
# Groups:   Region, Division [10]
   Region    Division           Urbanicity    EnergyCost EnergyCost_se
   <fct>     <fct>              <fct>              <dbl>         <dbl>
 1 Northeast New England        Urban Area          2629.          95.2
 2 Northeast New England        Urban Cluster       1889.         214. 
 3 Northeast New England        Rural               2865.         109. 
 4 Northeast Middle Atlantic    Urban Area          2133.          32.9
 5 Northeast Middle Atlantic    Urban Cluster       2259.         193. 
 6 Northeast Middle Atlantic    Rural               2420.         108. 
 7 Midwest   East North Central Urban Area          1632.          43.2
 8 Midwest   East North Central Urban Cluster       1654.          91.9
 9 Midwest   East North Central Rural               2263.          74.5
10 Midwest   West North Central Urban Area          1636.          71.1
# i 20 more rows
```

2. What is the median electric cost (DOLLAREL) for housing units in the South Region? What is the 95% confidence interval?

```
med_billsouth<-recs_des |>
  filter(Region=="South") |>
  summarize(
    MedElBill=survey_median(DOLLAREL,
                            vartype="ci")
  )
med_billsouth
```

```
# A tibble: 1 x 3
  MedElBill MedElBill_low MedElBill_upp
      <dbl>         <dbl>         <dbl>
1     1503.         1445.         1556.
```

The median electric cost for housing units in the South is $1,502 ($1,444, $1,555).

3. Test whether daytime winter and daytime summer temperatures of homes are set the same.

```
daytemp_ttest<-recs_des |>
  svyttest(design=_,
           formula = I(WinterTempDay-SummerTempDay)~0,
           na.rm = TRUE)
daytemp_ttest
```

```


    Design-based one-sample t-test

data:  I(WinterTempDay - SummerTempDay) ~ 0
t = -21.944, df = 94, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -2.401177 -2.002709
sample estimates:
     mean
-2.201943
```

On average housing units have set the temperature lower in the winter than the summer, p-value=$9.2356568 \times 10^{-39}$.

4. Test whether average electric bill (DOLLAREL) varies by region (Region).

```
m1 <- recs_des |>
  svyglm(design=_,
         formula=DOLLAREL~Region,
         na.action=na.omit)
summary(m1)
```

```
Call:
svyglm(design = recs_des, formula = DOLLAREL ~ Region, na.action = na.omit)

Survey design:
Called via srvyr

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    1345.84      34.18  39.376  < 2e-16 ***
RegionMidwest  -150.33      40.01  -3.757 0.000301 ***
RegionSouth     284.68      42.45   6.707 1.58e-09 ***
RegionWest     -200.14      47.28  -4.233 5.46e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 603803.4)

Number of Fisher Scoring iterations: 2
```

Yes, there is evidence that the average electric bill varies by region.

5. Fit a regression between the cooled square footage of a housing unit (TOTCSQFT) and the total amount spent on energy (TOTALDOL).

```
m2 <- recs_des |>
  svyglm(design=_,
         formula=TOTALDOL~TOTCSQFT,
         na.action=na.omit)
summary(m2)
```

```
Call:
svyglm(design = recs_des, formula = TOTALDOL ~ TOTCSQFT, na.action = na.omit)

Survey design:
Called via srvyr

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.378e+03  2.664e+01   51.71   <2e-16 ***
TOTCSQFT    3.503e-01  1.893e-02   18.50   <2e-16 ***
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 815669.9)

Number of Fisher Scoring iterations: 2
```

For each additional cooled square foot, the total energy cost increases by $0.35.